Selection of biological indicators appropriate for European soil monitoring

Stone, D; Ritz, K; Griffiths, BS; Orgiazzi, A; Creamer, RE

Published in:
Applied Soil Ecology

DOI:
10.1016/j.apsoil.2015.08.005

First published: 01/01/2016

Document Version
Peer reviewed version

Selection of biological indicators appropriate for European soil monitoring

D. Stone[1,2], K.Ritz[3,] B.G. Griffiths[4], A. Orgiazzi[5], R.E. Creamer[2]*

[1]Faculty of Biological Sciences, Leeds University, Leeds, UK, [2]Teagasc, Johnstown Castle Research Centre, Ireland, [3]Division of Agricultural & Environmental Sciences, University of Nottingham, Sutton Bonington Campus, Loughborough LE12 5RD, UK [4]Crop & Soils Systems Research Group, SRUC, Edinburgh, EH9 3JG, UK, [5] European Commission, Joint Research Centre (JRC), Institute for Environment and Sustainability (IES), Land Resource Management Unit, Via E. Fermi, 2749 – 21027 Ispra (VA), ITALY.

Corresponding author (rachel.creamer@teagasc.ie)

Abstract

The selection of biological indicators for monitoring progress towards policy goals for soil quality should be without bias and in line with individual scenarios of need. Here we describe the prescription of a suite of appropriate indicators for potential application in such monitoring schemes across Europe. We applied a structured framework of assessment and ranking (*viz.* a 'logical sieve'), building upon published data and a new survey taken from a wide section of the global soil biodiversity research and policy community.

The top ten indicators included four indicators of biodiversity (three microbial and one meso-faunal) by various methods of measurement, and three indicators of ecological function (Multiple enzyme assay, Multiple substrate-induced respiration profiling, and 'Functional genes by molecular biological means'). Within the techniques assessed, seven out of the top ten indicators made use of molecular methods.

Keywords (5)

Logical sieve, indicators, monitoring, soil biodiversity, soil ecosystem function

1. Introduction

National and international policy development regarding soil quality requires data on the current situation to create a baseline, or normal operating range, to which new policies can be applied. Enforcement of policy in turn requires monitoring, using pre-described indicators, to identify improvement or degradation (Turbe *et al.*, 2010). In 2001 the OECD laid down a set of criteria for agri-biodiversity indicators. These need to be measurable, based in sound science, able to be interpreted by policy makers, and policy relevant. They also need to allow monitoring at multiple spatial and temporal scales (OECD, 2001).

The majority of soil processes are intrinsically linked to soil biota, although specific relationships are complex and as yet not fully described in all cases (Ritz *et al*. 2009). However it is under these conditions of uncertainty that biological indicators are key in monitoring soil quality (here defined as the ability to deliver key soil processes) in ways that other indicators are not. By measuring organisms - which by being inherently adaptive - integrate multi-dimensional phenomena including time, it is possible at least in principle to ascertain the full potential of a soil to deliver key soil processes (Ritz *et al*. 2009).

However large-scale soil monitoring campaigns are expensive both in terms of labour and capital needed to acquire enough data for accurate conclusions. Selecting the best group of indicators is therefore important to reduce costs and provide data which is fit

for purpose, and can enable policy making and policy implementation in relation to soil quality into the future. This is however not as easy as it sounds. Previous attempts, for example the ENVASSO project working towards an operational set of EU-wide criteria and indicators to provide monitoring in Europe, found it difficult to arrive at a small set of indicators due to the complexity of soil biota and functions (Bispo *et al.* 2009). Faber *et al.* (2013), investigating the level of consensus in choice of indicator found that across Europe the use of different indicators varied from country to country, and where an indicator was used frequently (for example nematodes as a faunal indicator) this was driven by innate inclusion in a large scale project, rather than intrinsic potential of using an indicator alone.

Working to solve this problem, Ritz *et al.* (2009) developed a 'logical sieve' procedure to select indicators in a structured manner and without bias, and applied it to the context of soil monitoring at the specific scale of the UK. The logical sieve method formally considers the relation of a candidate indicator to a range of criteria based upon applicability, discrimination ability, and involvement in more than one function, plus a range of technical attributes associated with a method. It then invokes a series of sieving processes to exclude those not fit for purpose, and ranks the remaining indicators on a numerical basis. This expansion of selection criteria away from the more common technically-focused assessment (i.e. how easy or cheap is a specific method), gives the results of a logical sieve greater robustness but also provides an audit trail to the ranking that allows flexibility in adjusting parameters as needs or knowledge changes.

Of course, a single monitoring scheme cannot be a perfect fit to all scenarios. The logical sieve approach is advantageous, since the method is run on a specific scenario which has been laid out in advance, and the scenario can be changed and the process re-run without much additional effort. This provides different suites of indicators for different scenarios making the process flexible and more useful. Ritz *et al*. (2009) deployed the logical sieve on a scenario of a monitoring scheme for soil quality within the UK, incorporating the then contemporary state of methods developed for measuring soil biological indicators. Expanding on this paper, and working with the concepts and equations laid down for the logical sieve, we investigated whether a suite of indicators could be found which would enable a soil quality monitoring scheme to function over the whole of the European land area. This represents a more complex scenario than the one laid out in Ritz *et al*. (2009) by encompassing a wider range of climatic zones. The sieve itself was also updated to take into account developments in methods for measuring biological indicators of soil quality.

In the last decade with the rise of molecular techniques in particular, the suite of available biological indicators has changed, whilst at the same time there have been advances in knowledge of the functional roles of below-ground biodiversity. (Bardgett & van der Putten, 2014). This was foreshadowed by Ritz *et al*. (2009), but the pace of development has accelerated in recent years with massive advances in technology and the reduction in the costs of using molecular tools, which continue apace (Orgiazzi *et al.,* 2015). For example, with regard to next generation sequencing (NGS) three new platforms were released in 2011 alone: Ion Torrent's PGM Pacific Biosciences' RS and the Illumina MiSeq (Quail *et al*., 2012). Alongside these advances in technology are developments of bio-informatics tools that manage large data flows compare data

with specialised databases and extract relevant information thus opening up new perspectives for investigating the soil microbiome (Uroz *et al.*, 2013).

This work expands on in the list of indicators put forward by Faber *et al*. (2013), incorporates updates on logical sieve selection procedures and is then applied to a global-range survey of the soil biodiversity community in order to select indicators for future use in monitoring schemes.

2. Method

The logical sieve technique (Ritz *et al.*, 2009) was applied to the list of indicators originally described in Faber *et al.* (2013) and included here as Table 1. These indicators were shortlisted from the suite used in Ritz *et al.* (2009), using analysis of meta-data of soil biodiversity indicators used across Europe (Faber *et al*., 2013). Where indicators could be measured or assessed by more than one method, these were included as separate indicator/method combinations (Table 1). Given that genetic-based molecular methods are in a huge state of flux in terms of development (i.e. resolution, throughput, interpretation), indicators based on these approaches were sub-categorised in relation to the underlying objective of the approach. Fingerprinting methods (terminal restriction fragment length polymorphism (TRFLP) and automated ribosomal intergenic spacer analysis (ARISA) were included as an overall means of measuring microbial community structure. Pyrosequencing was included as a broad scale approach to soil biodiversity in general. While a molecular method for measuring the biodiversity of bacteria and archaea (which could include TRFLP, ARISA and pyrosequencing) was included as a specific determination on these microbial groups (Table 1).

A scenario was established to outline the purpose of the monitoring for which the indicators were being assessed. The logical sieve methodology was applied to establish key indicators for the assessment and potential monitoring of ecosystem functions specifically: (1) habitat for soil biodiversity; (2) C cycling and storage; (3) cycling of nitrogen (N) and phosphorus (P) across the whole productive non-urban land area of European member states for use in future potential monitoring.

Following on from the work of Faber *et al*. (2013) where indicators were scored against criteria taken from OECD (2002); UNEP (2007), Ritz *et al*. (2009); and Turbé *et al*. (2010) an online questionnaire was developed. The original questions used in Faber *et al*. (2013) were based on responses to a pilot run of the survey and communication with a small panel of experts in the field. The questionnaire used in this study expanded this initial study using additional criteria adapted from the logical sieve of Ritz *et al*. (2009) in order to provide a complete assessment of each indicator.

To facilitate the assessment of indicators by those completing the survey a set of choices were developed for answering each question. Each question had three possible answers consisting of the 'best possible' option a 'medium' option that was neither ideal nor problematic for the measurement of the indicator and a 'worst-case' option where the measurement of the indicator was affected. An additional answer of 'essentially no knowledge' was included for all questions allowing respondents to indicate that they considered that they did not know enough about a certain indicator to answer the question with sufficient confidence.

153 The final questionnaire contained 15 questions with associated answer choices (Table
154 2) and was html coded as a web survey by the Joint Research Centre (JRC) of the
155 European Commission. The questionnaire was made available online for three
156 months (April-June 2014) by JRC and was open to all who wished to participate. It
157 was advertised through the Global Soil Biodiversity Institute network
158 (http://www.globalsoilbiodiversity.org), EcoFINDERS project partners, and European
159 research scientists who work in the field of soil biodiversity and ecosystem function,
160 and related fields. JRC collated the results of the web survey and returned the raw
161 data whereupon it was amalgamated, checked, sorted and finally passed through the
162 equations of the logical sieve as per Ritz *et al*. (2009). These equations are
163 summarised in Figure 1.
164
165 Questions relating to skill requirements measurement equipment requirements and
166 costs were used to calculate the technical factor (F_T). The weighting factors (*WCi*)
167 used in the calculation of F_T were taken from Ritz *et al*., (2009) which were the result
168 of stakeholder consultation with the scientific community and likely end-user public-
169 bodies (reproduced here in Table 2).
170
171 Questions 12 and 14 (Table 3) were incorporated into the discrimination factor [S_D]
172 and taken forward into the applicability/discrimination factor [F_{AD}]. A soil function
173 factor for each indicator was calculated by the amalgamation of scores allocated by a
174 small panel of experts for each indicator. A final aggregated factor [F_A] was
175 calculated for each indicator.
176

3. Results

There were 61 respondents to the web survey. Three of the responses were rejected because the participants supplied non-usable answers (i.e. no knowledge). The responses of the remaining 58 were taken forward for analysis in the logical sieve.

3.1 Response description

The majority of indicator/question combinations received between 25 and 45 respondents out of a possible 58 (range 43-78%). Many participants used the 'no knowledge' option for one or more of the indicator/question combinations. No question was answered for all indicators by all participants.

The lowest response for an indicator/question combination was 14 respondents for a question on standardisation [Q5] using the indicator of 'Mites (by molecular techniques)'. The highest response was 51 respondents for a question regarding equipment needed [Q2] to measure the indicator of 'Respiration (using methods to measure basal respiration)'. Q2 and Q15 (understandability) were answered by the greatest proportion of respondents.

Molecular indicators generally had a poor response, overall with the lowest proportion with regard to questions on standardisation (Q5, 14, 15, 17, 19: all under 20 respondents).

Only two indicators returned 50 or more responses ('Respiration (all basal methods)' as above and 'Capital start up needed [Q6] for litter bags').

Indicators with a median response value under 30 included mainly molecular methods ('Enchytraeids', 'Mites', 'Collembola', and 'Protozoa') and 'ChipTechnology'. These indicators were also ranked lowly in the technical factor.

An average response was calculated per indicator/question combination. 'No knowledge' answers were removed from the dataset before averages were calculated.

3.2 Technical rankings

Technical Factor Scores [F_T] were calculated for each indicator using the sum of weighted responses to questions (Table 4). Measurements of soil respiration scored highest with molecular methods scoring poorly over the weighted responses to technical questions.

The top three indicators after ranking by F_T scores ('Basal respiration', 'Substrate induced respiration' and 'Nitrification potential') all scored highly on four particular questions. These four questions addressed labour in the lab (Q7: 'Does not require senior technical staff or an experienced researcher') analysis costs (Q8: 'Low cost per sample in terms of consumables') ease of sampling (Q11: 'A short one off sampling occasion in the field with a low volume of soil needed') and reproducibility of results (Q13: 'Reproducible results across different laboratories'). Though the top indicators scored more than 2 on a number of technical questions they all scored an average of more than 2.5 out of a possible 3 in these four specific areas.

There was no bias of either cost questions or method of measurement questions within the technical factor scores. Weighted technical scores were grouped by

225 'Measurement' (6 questions) and by 'Cost' (6 questions) and a single index value
226 produced (i.e. one measurement value as a % of the possible total weighted score and
227 one cost value as a % of the possible total weighted score). These two index values
228 were added together to produce a F_T score that had an equal contribution from
229 measurement and cost questions. The top ten indicators then remained essentially the
230 same. The top five indicators did not change in any way; indicators ranked 6 and 7
231 changed places ('Respiration measured by MSIR' and 'Litter bags' respectively);
232 indicators ranked 8 and 9 also switched places ('Collembola identified by
233 morphological methods' and 'Molecular microbial biomass' respectively).
234
235 3.3 Applicability and discrimination
236 Applicability/Discrimination Factor Scores [S_A] and [S_D] were calculated for each
237 indicator using scores for applicability taken from Ritz *et al*. (2009) and responses to
238 questions about discrimination ability of indicators (Table 5).
239
240 Molecular indicators of biodiversity and indicators of ecosystem function scored
241 highly in discrimination [S_D]. The top three indicators after ranking by
242 Discrimination Factor Scores [S_D] ('Multiple enzyme assay', 'PLFA', and 'Molecular
243 methods for assessing bacteria and archaea') all scored highly on overall sensitivity
244 (Q14: indicator is sensitive to more than one of either: land use disturbance or soil
245 type).
246
247 In terms of Applicability [S_A] 'Earthworms' and 'Enchytraeids' were both allocated
248 low S_A scores due to the fact that they are not ubiquitous to all European bio-
249 geographical zones.

250

251 3.4 Functionality

252 To assess the usefulness of the indicators in describing changes in functionality of soil
253 experts were asked to review the influence values taken from Ritz *et al*. (2009) for the
254 following functions: (1) Food and fibre production [S_{FF}]; (2) Environmental
255 interactions [S_{EI}]; (3) Habitat for biodiversity [S_{HB}]. The updated influence values
256 were developed taking into account new methods for analysis and new knowledge of
257 ecological importance of the indicators involved.

258

259 To test indicator rankings for all functions, scores for each function were multiplied
260 together to create a Function factor score [F_{SF}] (Table 6a). This sieved out indicators
261 ranked 'not pertinent' (0) for any of the three functions in order to assess the overall
262 potential across functions. In this ranking 'Potential nitrification' and 'Protista
263 diversity' scored poorly: 'Nitrification potential' due to low influence in the function
264 of habitat and biodiversity provision, and 'Protista diversity' due to low influence in
265 the function of S_{FF}.

266

267 The second ranking assumed that functions were equal and indicators did not need to
268 hold influence over *all* functions in order to score highly. In this ranking F_{SF} was
269 calculated from the average of the three functions (Table 6b). 'Nitrification potential'
270 scored better in this method of assessing functionality.

271

272 'Earthworms', 'Fungi', and 'Bacteria biodiversity' indicators scored highly in both.

273

274 3.5 Aggregated Factor Scores

F_{AD} was calculated from S_A, S_D and F_{SF} scores (using both multiplied and averaged [F_{SF}] values). The Aggregated factor [F_A] was then calculated for each indicator by multiplying F_{AD} by F_T (Figures 2 and 3 for multiplied and averaged [F_{SF}] values respectively).

As the Aggregated factor scores [F_A] were calculated from previous factors scores those indicators that performed poorly in previous rankings were found at the bottom of the F_A rankings. These included 'Earthworms' (ranked lowly in S_A – applicability) and 'Nitrification potential' and 'Protozoa biodiversity' (ranked lowly in F_{SF} – function).

The top ten indicators included four indicators of biodiversity (three microbial and one meso faunal) by various methods of measurement and three indicators of ecological function (multiple enzyme assay, multiple substrate induced respiration and the exploration of functional genes by molecular means). Within the techniques assessed, seven out of the top ten indicators made use of molecular methods.

Other soil fauna biodiversity indicators ('Nematodes' and 'Micro-arthropods') were found in the middle of the rankings, above the remaining indicators for ecological functions such as 'Biolog' and 'Basal respiration' (not ranked highly in F_{SF}) or 'Litter bags' and 'Bait lamina' (ranked lowly in S_D – discrimination).

4 Discussion

4.1 Response description

Any voluntary survey such as this contains the potential for self-selection bias and we acknowledge that respondents to the survey potentially had an interest in the outcome.

In part, the survey was designed to be self-selecting in order to access the knowledge and expertise of those specialising in the monitoring of soil biodiversity and soil ecosystem functioning. However, bearing this in mind, various measures were incorporated into the survey to reduce the effect of any potential bias.

Method assessment on a technical level was broken down into small discrete areas with clear and measurable criteria for assessment (cost of laboratory equipment, labour time in the field etc.). This reduced the number of qualitative responses and the potential for participants to over mark their favourite method.

Questions relating to sensitivity or reproducibility of methods will always ultimately be subjective. To address this, the design of the survey restricted participants to three categories of response with clear criterion guides for each category (plainly presented on the screen alongside the question), in order to prevent subtle over- or under-marking. To address blatant such marking, the survey was advertised to as broad a group of potential participants as possible in order to balance the bias between areas of research interest. The authors themselves had no bias to any particular group and advertised the survey amongst general scientific communities through mailing lists and fora in order to get a good range of technical backgrounds. In addition, though some indicators drew a less numerous response, by creating an average value per question per indicator, the potential for the number of respondents assessing an indicator to affect the overall value for that indicator was reduced.

It was clear from the answers collected by the web survey that scientists though knowledgeable in their field cannot be expected to have the information to answer questions about every selected indicator of soil biodiversity and ecosystem function. The use of the ‘no knowledge’ response at some point by all participants in some cases for all indicators except one confirmed that the incorporation of this option was

necessary and its general use is reassuring in the sense that such a return would be anticipated if respondents were being duly considerate in their returns. Without the option to not answer the question the survey would likely have been biased by guessed answers and not performed the function it was designed for.

It is however, noted that the potential for self-selection bias cannot be completely denied when acting on the conclusions drawn by this paper.

4.2 Technical rankings

Technical factor scores [F_T] give an indication of the practicality associated with the measurement of a particular indicator. The questions used for F_T scores are those most commonly assessed when choosing an indicator i.e. cost and difficulty of the measurement method. Indicators considered to have a high F_T are generally incorporated into monitoring schemes without further assessment. As this paper shows the indicators in the top 10 F_T scores were not always those which scored highest overall. This is an advantage of the logical sieve method. The top ten indicators after ranking by F_T scores are dominated by measurements of ecosystem function. Methods used in the measurement of ecosystem function are often simple and cost effective especially when compared to the high cost of molecular techniques and the high labour demand associated with soil fauna identification (Wu *et al*. 2009).

4.3 Applicability and discrimination

Applicability [S_A] tests the ubiquitous nature of each indicator and thus forms an intrinsic part of the scenario tested. In the scenario tested in this paper covering the whole productive non-urban land area of European member states, the restricted range of earthworms was a major factor, resulting in a low ranking overall despite high F_T

scores. Having evidence of why an indicator cannot be recommended for a monitoring scheme (in this case due to non-presence in some monitoring sites) gives the potential for the scheme to be adapted before implementation. This highlights another benefit of the logical sieve in that indicators are ranked and rejected for practical reasons and the sieve could always be re-run with changed parameters to adjust the indicators recommended in accord with the context (particularly spatial scale and range of biomes involved).

Discrimination [S_D] tests the sensitivity of each indicator to a wide range of environmental conditions. Five of the top ten indicators within this category were for ecological function (two for soil respiration which is an indicator for C cycling and storage). Different soils types and management schemes produce soils with different abilities to perform ecosystem functions. Indicators which measure ecosystem function are therefore intrinsically able to discriminate between these different soil conditions and are often recommended for monitoring schemes (Faber *et al.*, 2013).

Three indicators of biodiversity (two indicators of microbial diversity plus 'Nematode diversity') were placed in the top 10 rankings. A great deal of species data is collected when molecular methods to determine biodiversity are employed. This species data can be used to create a more detailed picture of the community present and with the massive niche diversity in soil habitats species diversity varies from scenario to scenario leading to high levels of discrimination. In addition if trait information is available species data can be linked to soils' ability to perform certain ecological functions leading to greater discrimination between them.

375 4.4 Functionality

376 Functionality [F_{SF}] changes the recommended indicators dependant on the function or
377 functions wanted, removing certain indicators from the logical sieve that might
378 otherwise be recommended. The example of the indicator 'Potential nitrification'
379 shows how vital it is to be clear about which functions should be included in the
380 scenario used for each logical sieve process. When the scenario was that *all* functions
381 needed to be monitored by the indicators selected, potential nitrification scored poorly
382 despite being reasonably ranked in all other factors. However, when the scenario was
383 changed to a looser monitoring of three functions but each indicator was not required
384 to be able to monitor all three, 'Potential nitrification' scored higher in the ranking.

385

386 If the criteria for selection requires indicators necessarily that are ubiquitous in terms
387 of function then it is the molecular methods for determining the biodiversity of the
388 microbial community that score most highly. If the guidelines are a little looser,
389 allowing the inclusion of indicators that are relevant to less than all of the functions
390 required then methods of measuring ecological function rise up the rankings.

391

392 4.5 Aggregated factor scores

393 The top ten indicators after the full suite of sieves was applied ([F_A] scores) were a
394 mix of measures of soil biodiversity and soil ecosystem function. When re-examining
395 the sieving processes it becomes clear that this is due to the discrimination potential
396 and applicability to a range of functions for biodiversity indicators, and a mix of
397 technical factors and discrimination potential for indicators of soil ecosystem function
398 provision. The top ten rankings were dominated by molecular methods of measuring
399 both biodiversity and ecological function (seven indicator/method combinations).

400 This is interesting, bearing in mind that in terms of responses to the questionnaire,
401 molecular indicators generally had a poor response overall when compared to the
402 number of respondents who answered questions on other indicators. The results of
403 this suggest that despite rapid advances in molecular techniques over the last decade,
404 the potential of molecular methods has not yet percolated out to the majority of the
405 scientific community surveyed in this questionnaire.
406
407 Within the top indicators for biodiversity, both 'Bacteria and Archaea diversity' and
408 'Fungi diversity' were only in the middle rankings for Technical factors [F_T] but
409 scored in the top five for Discrimination potential [S_D] and high for relevance to
410 function [F_{SF}] ('Bacteria and Archaea' were number one). 'Mite biodiversity' scored
411 poorly in [F_T] and was only in the middle of the table for [S_D], but scored in the top
412 ten for relevance to function [F_{SF}]. Though methods of determining biodiversity are
413 often time consuming and/or costly (Lawton *et al*., 1998, Andre *et al*., 2001 with
414 regard to 'scientist hours' needed for morphological identification), it can be seen
415 from the results of this exercise that the potential of these indicators in terms of
416 discrimination and relevance to more than one function results in them being
417 recommended for use as indicators for monitoring schemes.
418
419 Both 'Respiration (Multiple Substrate Induced Respiration)' and 'Molecular microbial
420 biomass' were in the top ten for Technical rankings [F_T] and the top ten for
421 Discrimination potential [S_D]. 'Functional Genes' and 'Multiple enzyme assay' were
422 mid-range in Technical rankings [F_T] but in the top five for Discrimination potential
423 [S_D]. 'Functional genes' also scored within the top five for Function factor scores
424 [F_{SF}]. With regard to indicators for ecological function, methods such as 'MSIR' or

425 'Molecular microbial biomass', which scored well in ease of use and discrimination
426 potential, are restricted in the number and variety of functions to which they are
427 relevant and are therefore not ranked in the top five. Whereas 'Functional genes' and
428 'Multiple enzyme assays', though not ranked as highly in Technical rankings [F_T],
429 scored highly both in discrimination potential and in relevance to more than one
430 function. This resulted in these two indicators being ranked higher in the overall
431 standings.
432
433 When not using the logical sieve, it is often indicators that score highly in Technical
434 factor [F_T] criteria (those considered the easiest to measure) that are selected for
435 monitoring schemes. However, as seen in this paper, when comparing indicators with
436 high F_T scores as opposed to those that were still ranked highly at the end of the
437 sieving process, this can produce a decision making process that fails to take into
438 account low discrimination ability of an indicator or does not recognise that an
439 indicator may not be present across the full range of environments found within the
440 monitoring scheme.
441
442 The functions selected within the scenario, in this case soil biodiversity habitat, C
443 cycling and nutrient cycling, have an intrinsic effect on the highest ranking indicators.
444 This allows the logical sieve method to be precisely calibrated for each monitoring
445 scheme to provide scheme specific indicators.
446
447 **5 Conclusion**
448 For the scenario provided, i.e. the assessment and potential monitoring of ecosystem
449 functions specifically: (1) habitat for soil biodiversity; (2) C cycling and storage; (3)

cycling of nitrogen (N) and phosphorus (P) across the whole productive non-urban land area of European member states, the top ranking biological indicator was 'Bacteria and archaea diversity measured by molecular methods'.

The top ten indicators included three related to biodiversity ('Bacteria and archaea', 'Fungi' and 'Mites') and four related to ecological function ('Functional genes', Multiple enzyme assay', 'Respiration (multiple substrate induced respiration)' and 'Molecular microbial biomass'). The list was dominated by molecular methods to measure both biodiversity and ecological function (7 indicator/method combinations).

These selected indicators are recommended only under advisement. As shown in this paper, it is only when applying the full suite of logical sieves that the indicators selected can be termed fit for monitoring the (precise) scenario intended. Careful prescription of the scenario is also important as the scale of the monitoring programme, the level of discrimination needed, and the need for pertinence to more than one function, all have the potential to change the ranking of indicators.

Acknowledgements

This work was supported by the European Union within the projects EcoFINDERS (FP7-264465), the Joint Research Council (JRC), the Scottish Government Rural and Environment Science and Analytical Services Division Theme 3 Land Use. We sincerely thank the individual respondents to the survey questionnaires, and members of the Joint Research Centre, and the Global Soil Biodiversity Initiative, for their assistance and engagement.

475 **References**

476 Andre, H.M., Ducarme, X., Anderson, J.M., Crossley, D.A., Koehler, H.H., Paoletti,
477 M.G., Walter, D.E., Lebrun, P., 2001. Skilled eyes are needed to go on studying the
478 richness of the soil. Nature. 409, 761.

479

480 Bardgett, R.D., & van der Putten, W.H., 2014. Belowground biodiversity and
481 ecosystem functioning. Nature. 515, 505–511

482

483 Bispo, A., Cluzeau, D., Creamer, R., Dombos, M., Graefe, U., Krogh, P.H., Sousa,
484 J.P., Peres, G., Rutgers, M., Winding, A., and Rombke, J., 2009. Indicators for
485 monitoring soil biodiversity. International Environmental Assessment Policy and
486 Management. 5(4), 717-719

487

488 Faber, J.H., Creamer, R.E., Mulder, C., Römbke, J., Rutgers, M., Sousa, J.S., Stone,
489 D., Griffiths, B.S., 2013. The Practicalities and Pitfalls of Establishing a
490 Policy-Relevant and Cost-Effective Soil Biological Monitoring Scheme. Integrated
491 Environmental Assessment and Management. 9(2), 276–284

492

493 Lawton, J.H., Bignell, D.E., Bolton, B., Bloemers, G.F., Eggleton, P., Hammond,
494 P.M., Hodda, M., Holt, R.D., Larsenk, T.B., Mawdsley, N.A., Stork, N.E., Srivastava,
495 D.S., Watt, A.D., 1998. Biodiversity inventories, indicator taxa and effects of habitat
496 modification in tropical forest. Nature. 391, 72-76.

497

498 Morvan, X., Saby, N.P.A., Arrouays, D., Le Bas, C., Jones, R.J.A., Verheijen, F.G.A.,
499 Bellamy, P.H., Stephens, M., Kibblewhite, M.G., 2008. Soil monitoring in Europe: A

review of existing systems and requirements for harmonisation. Science of The Total Environment. 391(1), 1-12

[OECD] Organization for Economic Co-Operation and Development, 2002. Report on the OECD expert meeting on agri-biodiversity indicators. November 2001: Summary and recommendations. OECD Joint Working Party on Agriculture and Environment COM/AGR/CA/ENV/EPOC (2002) 35. Paris: OECD.

Orgiazzi, A., Dunbar, M.B., Panagos, P., de Groot, G.A., Lemanceau, P., 2015. Soil biodiversity and DNA barcodes: opportunities and challenges. Soil Biology and Biochemistry. 80, 244-250

Quail, M.A., Smith, M., Coupland, P., Otto, T.D., Harris, S.R., Connor, T.R., Bertoni, A., Swerdlow, H.P., and Gu, Y., 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics. 13, 341

Ritz, K., Black, H.I.J., Campbell, C.D., Harris, J.A., Wood, C., 2009. Selecting biological indicators for monitoring soils: A framework for balancing scientific and technical opinion to assist policy development. Ecological Indicators. 9, 1212-1221

Turbé A., De Toni, A., Benito, P., Lavelle, P., Ruiz, N., Van der Putten, W.H., Labouze, E., Mudgal, S., 2010. Soil biodiversity: Functions, threats and tools for policy makers. Bio Intelligence Service, IRD, and NIOO, Report for European Commission (DG Environment).

525

526 UNEP/CBD (2007). In-depth review of the programme of work on agricultural
527 biodiversity – Note by the Executive Secretary., UNEP.

528

529 Uroz, S., Ioannidis , P., Lengelle, J., Cébron , A., Morin, E., Buée, M., Martin, F.,
530 2013. Functional Assays and Metagenomic Analyses Reveals Differences between the
531 Microbial Communities Inhabiting the Soil Horizons of a Norway Spruce Plantation.
532 *PLOS* *ONE*.
533 http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0055929

534

535 Wu, T., Ayres, E., Li, G., Bardgett, R.D., Wall, D.H., Garey, J.R., 2009. Molecular
536 profiling of soil animal diversity in natural ecosystems: Incongruence of molecular
537 and morphological results. Soil Biology & Biochemistry. 41, 849–857

538 Table 1. List of indicators processed by logical sieve and their importance to either
539 Biodiversity or Ecosystem Function.

Indicator	Biodiversity	Function
Macro and Meso faunal diversity		
Earthworms (Morphological identification or Molecular methods)	X	X
Enchytraeids (Morphological identification or Molecular methods)	X	X
Mites (Morphological identification or Molecular methods)	X	X
Collembola (Morphological identification or Molecular methods)	X	X
Nematodes (Morphological identification or Molecular methods)	X	X
Protista diversity (Morphological identification or Molecular methods)	X	X
Microfaunal diversity		
Bacteria and Archaea species by molecular methods	X	
Fungi species (Morphological identification or Molecular methods),	X	
Bacteria & Fungi diversity through fingerprint methods (TRFLP, ARISA)*	X	
Pyrosequencing, of soil DNA**	X	X
PLFA,	X	
Ecosystem Function performed by soil biology		
Functional Genes (targeting antibiotic producers, nitrifiers, denitrifiers)		X
ChipTechnology (up regulation or down regulation of specific genes related to Ecosystem Function)		X
Molecular microbial biomass		X
Respiration (All basal methods),		X
Respiration (SIR-Glucose),		X
Respiration (Multiple Substrate Induced Respiration),		X
Respiration (Biolog)		X
Nitrification potential		X
Multiple Enzyme Assay,		X
Bait Lamina,		X
Litter Bags		X

540 *As a method of assessing microbial community structure

541 **Broad-spectrum pyrosequencing of whole soil DNA

542

Table 2. The weighting factors (WCi) used in the calculation of F_T, taken from Ritz *et al*., (2009) using the question: "What *weighting* would you assign to the criterion when considering a trans-UK (cross-habitat) measuring and monitoring programme? Weight from 0 (i.e. dismiss entirely) to 4 (maximum relative weight)."

	CATEGORY:	Weight	Question in sieve
1	THROUGHPUT: *How important is it to be able to have a high level of throughput (i.e. 100's per week) e.g.* *0 = not important - dismiss* *1 = relevant but not essential* *2 = valuable but not essential* *3 = valuable and preferred* *4 = vital!*	3	Q9
2	STORAGE: *How important is it to be able to store samples until they can be analysed, for up to 2 weeks post sampling?*	3	Not used
3	ARCHIVABILITY*: How important do you consider archiving of samples (or analytical products) e.g. for future monitoring comparisons or for currently unknown analyses to answer new questions?*	2	Q3
4	SAMPLE COLLECTION: *Does it matter that the site would need to be visited more than once for a particular method to get the data?*	3	Q11
5	HOW MUCH SOIL*: Smaller soil samples cost less, easier to sample and handle etc; is a smaller sample preferred?*	2	Q11
6	COST – HARDWARE: *Does it matter how much it costs, in terms of hardware, to analyse the soil?*	1	Q6
7	COST- LABOUR: *Does it matter how much it costs, in terms of people, to analyse the soil?*	3	Q7, Q9
8	EASE OF USE: *Is it important that the method is relatively easy to carry out?*	2	Not used
9	POTENTIAL REFERENCE MATERIAL: *How important is quality control (QC) via reference material?*	2	Q4
10	REPRODUCIBILITY OF RESULTS: *How much do you care about being able to reproduce the same results time after time?*	4	Q1, Q5, Q13
11	READY-TO-USE DEPLOYMENT STATUS: *Is it important that the method is well established and has standard operating procedures?*	0	Q5
12	INTERNATIONAL COMPARISONS: *If the method is used in soil monitoring schemes elsewhere, is this important for UK soil monitoring?*	2	Not used
13	UK INFRASTRUCTURE: *Is it important that we have the capacity at present to deliver this method?*	3	Not used

549 Table 3. Questionnaire questions and answer choices.

Question	Best Option	Middle Option	Worst Option
Q1. Skills What is the amenability of the method to ready application via a standard operating procedure when presented to a competent technician? Does it include a training element? (Ritz *et al*., 2009)	Straightforward No more than 1 day of training is required to carry out the analysis	Moderate A PhD student or skilled technician can carry out the analysis after no more than 1 week of training.	Specialised A fully trained specialist is required full time to carry out the analysis
Q2. Equipment Are you likely to find the required equipment in all labs (i.e. pH meter) or very few (confocal microscope).	All laboratories All laboratories can perform this method. Also applies to methods carried out in the field.	Standard medium level laboratories The method can be successfully performed in a mid-range laboratory	Very few, specialised laboratories The method can only be carried out in a specialised laboratory designed for that purpose
Q3. Archive What is the potential for archiving samples (i.e. over decades) in order to accurately re-determine these properties? (Ritz *et al*., 2009)	Air dried soil/ Fauna stored in Alcohol/Formalin Samples can be archived as air dried soil or, if fauna, in alcohol/formalin.	Freeze dried/-80 freezer Samples are only able to be archived for any length of time either freeze-dried, in a -20 degree freezer or in a -80 degree freezer.	Not able to be archived Samples cannot be archived for the length of time described.
Q4. Potential reference material Is the method amenable to the prescription and provision of reference material (Ritz *et al*., 2009)	Readily available materials exist Reference materials are available and affordable for most mid-range laboratories OR "in house" QCs are possible to use.	Yes, but materials are expensive/difficult to obtain Materials are expensive (cost more than 500 Euros)	No The method does not allow the use of a reference material.
Q5. Method Standardised Is the method standardised (i.e. is there an ISO/SOP available?	ISO available An ISO is available and used by all.	Peer reviewed SOP in the literature A clear and detailed SOP is available in the literature and generally followed by peers	Local lab SOP The method follows an SOP that is only described within the local laboratory
Q6. Capital Start-up What level of finance is required to equip a laboratory to perform this analysis? (Instrument cost only, not salary or	Low Less than 2000 Euros	Moderate Up to 50,000 Euros	Very Expensive More than 50,000 Euro

training)* *Additional notes:* *In DNA sequencing methods, if you do your sequencing "in-house" then the cost of sequencing equipment should be included when answering this question. If sequencing is contracted out, this will be covered in the later question "cost per sample" (Q8).*			
Q7. Labour requirement What are the human resource costs to realise the method and initial interpretation (including consideration of skill level required and associated salary). (Ritz *et al*., 2009)	<u>Low cost</u> Technical staff at a salary of 25,000 Euros per year	<u>Moderately expensive</u> Senior Technical staff or Experienced Researcher at a salary of 35,000 Euros per year	<u>Very Expensive</u> Senior Researcher or Specialist at a salary of 50,000 Euros per year
Q8. Cost per sample (consumables) What are the consumable resource costs to realise the method and initial interpretation. (Ritz *et al*., 2009) Cost per average sample. If cost varies with number of individuals per sample, then the cost of a mid-range sample should be recorded. *Additional notes:* *In DNA sequencing methods, if you do not do your sequencing "in-house" then the cost of contracting out sequencing should be included when answering this question.* *Cost per sample in terms of labour is addressed in the next question (Q9).*	<u>Low cost</u> Less than 2 Euros per sample	<u>Moderately expensive</u> Less than 20 Euros per sample	<u>Very Expensive</u> More than 20 Euros per sample
Q9. Labour intensity (laboratory) How long does it take to carry out the method from start to finish, up to the point of production of initial data results? Does not include pre-processing of sample or post processing of data.	<u>Less than 1 day</u> The method requires less than the full time labour of one person for one day.	<u>1 day to 1 week</u> The method requires either the full time labour of one person for more than one day OR takes more than one full day due to the length of time of some steps	<u>More than 1 week</u> The method requires either the full time labour of one person for over one week OR takes more than one week due to the length of time of some steps.
Q10. Labour intensity (pre-processing) How much pre-processing does the sample require before analysis (from field sample)?	<u><1 hour</u> Sample processing requires less than the full	<u>1 day</u> Sample processing requires either the full time	<u>1 week</u> Sample processing requires either the full time labour of

	time labour of one person for one hour.	labour of one person for one day OR takes one full day due to the length of time of some steps.	one person for one week OR takes one week due to the length of time of some steps.
Q11. Labour intensity (field) How long does field sampling take? Consider: Is one stop sampling in the field tenable? What mass of soil is needed for sampling and determination?	Low One- off sampling required AND less than 500 g needed from the field site AND sampling time less than one hour.	Medium Either one-off sampling required, but this sampling is a composite of a large area OR < 2.5 kg of sample is required OR sampling will take more than three hours.	High Either timelapse sampling is required, therefore multiple visits over short time frame (< 1 month) OR large amounts of sample are required (>5 kg), OR sampling time in total (if multiple visits are required) will take more than six hours.
Q12. Field Variability What is the inherent spatial variability of the parameter measured under average field conditions?	Low A single sample per site is adequate. This category includes the taking of a single composite sample per site.	Moderate Three to five individual field replicates are required within a single site to account for field variability. This does not apply to composite samples.	High More than five individual field replicates are required within a single site to account for field variability. This does not apply to composite samples.
Q13. Reproducibility of results What is the inherent ability for the method to generate reproducible results, given that full quality-control protocols are available and applied, including (assumed) availability of reference material.	High Reproducible results are expected across different laboratories when following an SOP.	Moderate Validation through an inter laboratory comparison is required for reproducibility of results.	Inherently poor The method is so sensitive to operator controlled conditions that it cannot be reproduced across laboratories
Q14. Overall sensitivity Is this indicator sensitive enough to be of use in monitoring programmes?	High sensitivity The indicator is sensitive to more than one of either: land use, disturbance or soil type.	Moderate sensitivity The indicator is sensitive to one of either: land use, disturbance or soil type.	Poor sensitivity The indicator is not sensitive to either: land use, disturbance or soil type OR the indicator is too sensitive to one or more of these parameters such that it

			becomes a problem when undertaking monitoring programmes OR there is not enough data available in the literature to make a conclusion.
Q15. Understandability How understandable is this indicator in the area of soil monitoring for soil health and biodiversity?	Public This indicator can be easily used to explain issues of soil health and biodiversity to the general public.	Policy makers/Land managers This indicator can be used to explain monitoring of soil health and biodiversity to policy makers. And can be used by them in future soil legislation and policy documents.	Scientists This indicator can be used to explain monitoring of soil health and biodiversity only to other scientists within the general fields of ecology, soil science, biology, and environmental chemistry.

550 Table 4. Ranking of indicators by Technical Factor Scores [F_T]. (Scores are out of a
551 potential total of 93)

Indicator	F_T
Respiration (All basal methods),	73.5
Respiration (SIR-Glucose),	73.2
Nitrification,	72.8
Earthworms-Morphology,	72.7
Biolog,	71.1
Litter Bags	70.3
Respiration (Multiple Substrate Induced Respiration),	69.9
Collembola-Morphology,	69.8
Molecular microbial biomass,	69.6
Bait Lamina,	69.2
Enchytraeids-Morphology,	68.9
Multiple Enzyme Assay,	68.4
Nematodes-Morphology,	68.1
Mites-Morphology,	67.7
PLFA,	67.2
Bacteria and Archaea-Molecular,	67.2
Fungi-Morphology,	65.8
Fungi-Molecular,	65.8
Bacteria & Fungi-fingerprints (TRFLP, ARISA,...),	65.3
Nematodes-Molecular,	64.3
Functional Genes (targetting antibiotic producers, nitrifiers, denitrifiers)	64.3
Collembola-Molecular,	63.8
Mites-Molecular,	63.5
Protozoa-Molecular, Bacteria and Archaea-Molecular,	63.4
Enchytraeids-Molecular,	63.4
Protozoa-Morphology,	62.4
Earthworms-Molecular,	60.7
ChipTechnology,	57.6
Pyrosequencing,	57.6

552

553 Table 5. Applicability/Discrimination Factor Scores [S_A] and [S_D]

Indicator	S_D
Multiple Enzyme Assay,	5.0
PLFA,	5.0
Bacteria and Archaea-Molecular,	4.7
Nitrification,	4.7
Fungi-Molecular,	4.7
Respiration (Multiple Substrate Induced Respiration),	4.6
Molecular microbial biomass,	4.6
Functional Genes (targetting antibiotic producers, nitrifiers, denitrifiers)	4.6
Nematodes-Molecular,	4.6
Respiration (SIR-Glucose),	4.5
Biolog,	4.5
Bacteria & Fungi-fingerprints (TRFLP, ARISA,...),	4.5
Mites-Molecular,	4.5
Pyrosequencing,	4.4
Fungi-Morphology,	4.3
Protozoa-Molecular,	4.3
ChipTechnology,	4.3
Nematodes-Morphology,	4.3
Respiration (All basal methods),	4.3
Collembola-Molecular,	4.2
Enchytraeids-Molecular,	4.0
Mites-Morphology,	4.0
Earthworms-Molecular,	3.8
Protozoa-Morphology,	3.8
Collembola-Morphology,	3.8
Earthworms-Morphology,	3.7
Enchytraeids-Morphology,	3.7
Litter Bags	3.7
Bait Lamina,	3.5

Indicator	S_A †
Earthworms-Morphology, Earthworms-Molecular, Enchytraeids-Morphology, Enchytraeids-Molecular,	0.1
Macro and Meso faunal diversity	
Mites-Morphology, Mites-Molecular, Collembola-Morphology, Collembola-Molecular Nematodes-Morphology, Nematodes-Molecular Protozoa-Morphology, Protozoa-Molecular	1
Microfaunal diversity	
Bacteria and Archaea-Molecular Fungi-Morphology, Fungi-Molecular Functional Genes (targetting antibiotic producers, nitrifiers, denitrifiers) Bacteria & Fungi-fingerprints (TRFLP, ARISA,...) Pyrosequencing, ChipTechnology PLFA	1
Ecosystem Function performed by soil biology	
Molecular microbial biomass Respiration (All basal methods), Respiration (SIR-Glucose) Respiration (Multiple Substrate Induced Respiration), Biolog, Nitrification Multiple Enzyme Assay	1
Bait Lamina, Litter Bags	

554 †Value of 0.1 used for non-ubiquitous species rather than 0 as in Ritz *et al*. (2009) in

555 order not to lose these indicators from the sieve whilst indicating their problems.

556 Table 6a. Function Factor Scores [F_{SF}] (Food and Fibre x Environmental interactions
557 x Habitat and Biodiversity provision) Total: 2 x 2 x 2 = 8

Indicator	**F_{SF}**
Bacteria and Archaea-Molecular,	8.00
ChipTechnology,	8.00
Earthworms-Molecular,	8.00
Earthworms-Morphology,	8.00
Functional Genes (targetting antibiotic producers, nitrifiers, denitrifiers)	8.00
Fungi-Molecular,	8.00
Fungi-Morphology,	8.00
Mites-Molecular,	8.00
Pyrosequencing,	8.00
Collembola-Molecular,	4.00
Collembola-Morphology,	4.00
Mites-Morphology,	4.00
Molecular microbial biomass,	4.00
Multiple Enzyme Assay,	4.00
Nematodes-Molecular,	4.00
Nematodes-Morphology,	4.00
Respiration (Multiple Substrate Induced Respiration),	4.00
Bait Lamina,	2.00
Biolog,	2.00
Enchytraeids-Molecular,	2.00
Enchytraeids-Morphology,	2.00
Litter Bags	2.00
Respiration (All basal methods),	2.00
Respiration (SIR-Glucose),	2.00
Bacteria & Fungi-fingerprints (TRFLP, ARISA,...),	1.00
PLFA,	1.00
Nitrification,	0.00
Protozoa-Molecular,	0.00
Protozoa-Morphology,	0.00

Table 6b. Function Factor Scores [F_{SF}] (Average of Food and Fibre, Environmental interactions, Habitat and Biodiversity provision) Total: (2 + 2 +2)/3 = 2

Indicator	F_{SF}
Bacteria and Archaea-Molecular,	2.00
ChipTechnology,	2.00
Earthworms-Molecular,	2.00
Earthworms-Morphology,	2.00
Functional Genes (targetting antibiotic producers, nitrifiers, denitrifiers)	2.00
Fungi-Molecular,	2.00
Fungi-Morphology,	2.00
Mites-Molecular,	2.00
Pyrosequencing,	2.00
Collembola-Molecular,	1.67
Collembola-Morphology,	1.67
Mites-Morphology,	1.67
Molecular microbial biomass,	1.67
Multiple Enzyme Assay,	1.67
Nematodes-Molecular,	1.67
Nematodes-Morphology,	1.67
Respiration (Multiple Substrate Induced Respiration),	1.67
Nitrification,	1.33
Bait Lamina,	1.33
Biolog,	1.33
Enchytraeids-Molecular,	1.33
Enchytraeids-Morphology,	1.33
Litter Bags	1.33
Respiration (All basal methods),	1.33
Respiration (SIR-Glucose),	1.33
Bacteria & Fungi-fingerprints (TRFLP, ARISA,...),	1.00
PLFA,	1.00
Protozoa-Molecular,	0.67
Protozoa-Morphology,	0.67

Figure 1. Equations of the logical sieve as per Ritz *et al*. (2009).

Figure 2. Aggregated Factor Scores [F_A] for the list of indicators assessed. [F_A] is calculated using *multiplied* Function Factor Score [F_{SF}] values and incorporates the ability of each indicator to be relevant to all the functions intended to be monitored.

Figure 3. Aggregated Factor Scores [F_A] for the list of indicators assessed. [F_A] is calculated using *averaged* Function Factor Score [F_{SF}] values and assess indicators in a scenario where they do not need to be relevant to all the functions intended to be monitored.

574 Figure 1.
575
576 $$F_A = F_{AD} \times F_T \quad (1)$$

577 Where

$$FT = \sum_{i=1}^{n} (S_{Ci} \times W_{Ci}) \ \ldots + \ \ldots (S_{Cn} \times W_{Cn}) \quad (2)$$

578

579 Where FT is the technical factor; S_{Ci} is the individual score for the indicator/question
580 combination *i*; W_{Ci} is the weighting value for the individual indicator/question
581 combination *i*; and *n* = number of indicator/question combinations.

582

583 And

584 $$FAD = S_A \times S_D \times F_{SF} \quad (3)$$

585 Where F_{AD} is the applicability/discrimination factor; S_A is the score for applicability,
586 taken from Ritz *et al*. (2009); S_D is the score for discrimination taken from the
587 averaged answer for questions 12 and 14; and F_{SF} is the soil function factor,
588 amalgamated from scorings given by a small panel of EcoFINDERS experts for each
589 indicator.

590

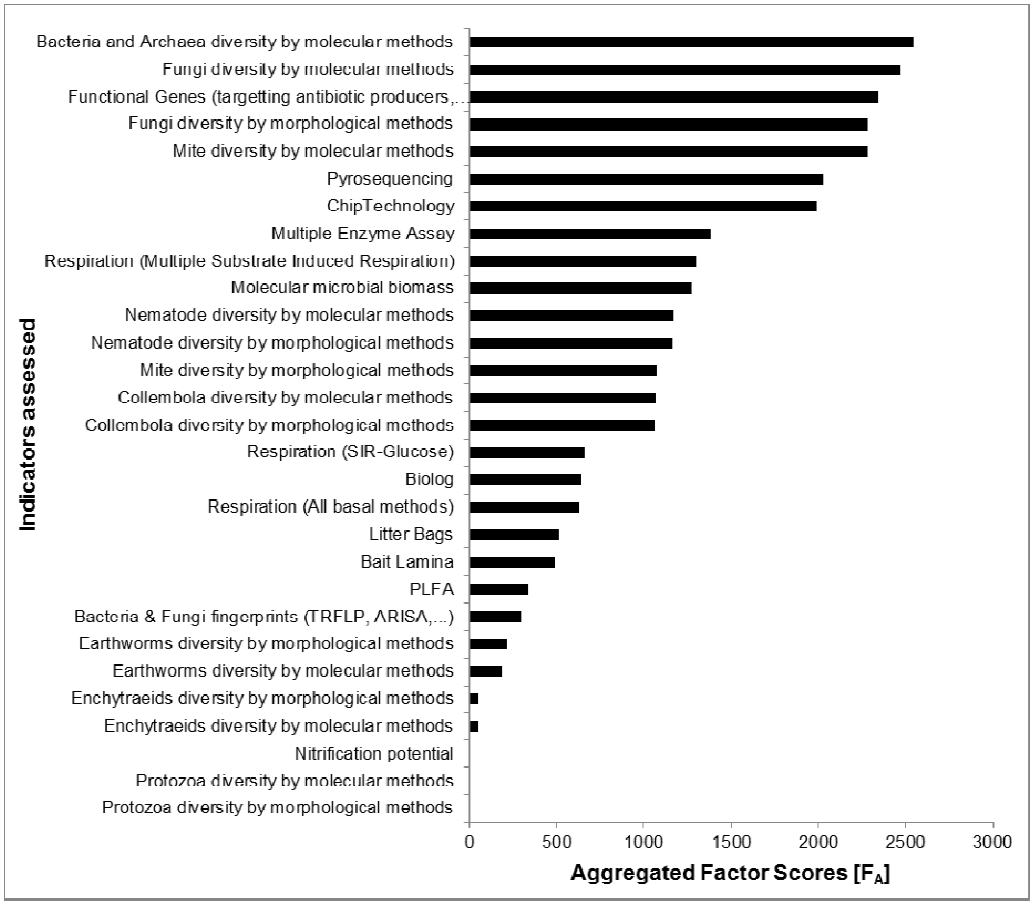
591 And

592 $$F_{SF} = S_{FF} \times S_{EI} \times S_{HB} \quad (4)$$

593 Where F_{SF} is the soil function factor; S_{FF} is the indicator score for the ecosystem
594 function of food and fibre production; S_{EI} is the indicator score for the ecosystem
595 function of environmental interaction; and S_{HB} is the indicator score for the ecosystem
596 function of promotion of habitat and biodiversity. The scores for each indicator for

these soil functions were amalgamated from scorings given by a small panel of EcoFINDERS experts for each indicator.

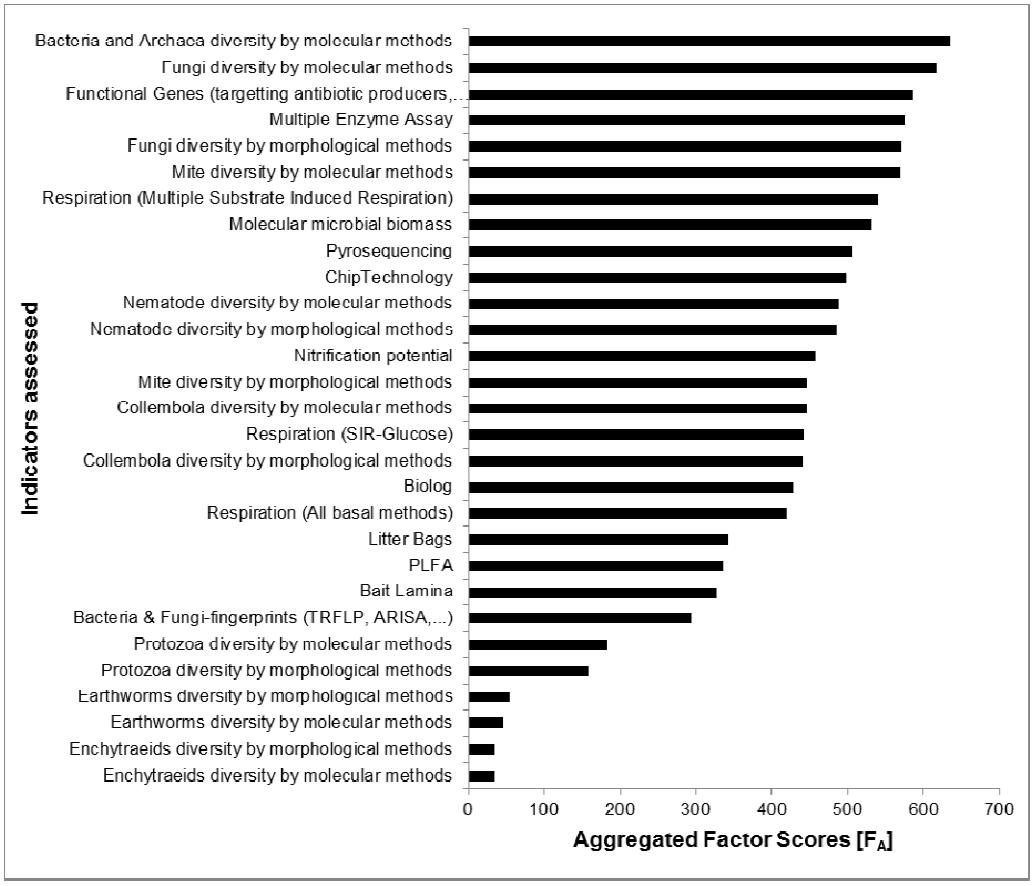
600 Figure 2



601

602

603 Figure 3



604